



AI-12655
JFW
C

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Eliot M. Case

Serial No.: 09/818,172

Filed: March 27, 2001

For: SYSTEM AND METHOD FOR CONVERTING TEXT-TO-VOICE

Attorney Docket No.: 1814 (USW 0617 PUS)

Group Art Unit: 2655

Examiner: Dmitry Brant

APPEAL BRIEF

Mail Stop Appeal Brief - Patents
Commissioner for Patents
U.S. Patent & Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This is an appeal brief from the final rejection of claims 1-11 of the Office Action dated June 15, 2004. This application was filed on March 27, 2001.

09/24/2004 SDIRETA1 00000054 09818172

01 FC:1402

330.00 OP

I. REAL PARTY IN INTEREST

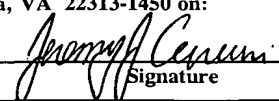
The real party in interest is Qwest Communications International Inc., a corporation organized and existing under the laws of the state of Delaware, and having a place of business at 1801 California Street; 38th Floor; Denver, Colorado 80202, as set forth in the assignment recorded in the U.S. Patent and Trademark Office on July 25, 2001 at Reel 012081/Frame 0320.

CERTIFICATE OF MAILING UNDER 37 C.F.R. § 1.8

I hereby certify that this paper, including all enclosures referred to herein, is being deposited with the United States Postal Service as first-class mail, postage pre-paid, in an envelope addressed to: Mail Stop Appeal Brief - Patents, Commissioner for Patents, U.S. Patent & Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450 on:

September 22, 2004
Date of Deposit

Jeremy J. Curcuri
Name of Person Signing


Signature

II. RELATED APPEALS AND INTERFERENCES

There are no appeals or interferences known to appellant, the appellant's legal representative, or assignee which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

III. STATUS OF CLAIMS

Claims 1-11 are pending in this application. Claims 1-11 have been rejected and are the subject of this appeal.

IV. STATUS OF AMENDMENTS

No amendment after final rejection has been filed.

V. SUMMARY OF CLAIMED SUBJECT MATTER

Claim 1 is the sole independent claim involved in the appeal. Claim 1 recites a method for converting text to concatenated voice by utilizing a digital voice library 12 (Figure 1) and a set of playback rules. The digital voice library 12 includes a plurality of speech items including words and syllables. The digital voice library 12 further includes a corresponding plurality of voice recordings. Each speech item corresponds to at least one available voice recording. The method comprises training the digital voice library to associate each syllable speech item with a literal text syllable of the particular syllable speech item (Figures 6 and 7).

This claimed subject matter is summarized in the application specification at page 1, line 19 - page 2, line 7. In more detail, Figure 1 depicts a digital voice library 12.

Digital voice library 12 is trained to associate each syllable speech item with a literal text syllable of the particular speech item. Details of the preferred embodiment are described in the specification at page 26, line 27 - page 28, line 15 and Figures 6 and 7. Figures 6 and 7 depict syllable-level conversion of text input as known words or literally spelled by syllable to spoken output as pre-recorded words or phonetically spelled by syllable. In this way, mappings of literal spellings to phonetic pronunciations of syllables (the training recited by claim 1) can then be used as the lookup criteria to select recordings of syllables for a syllable level concatenated speech output. Specification, page 27, lines 8-10.

According to the invention, literal spellings of syllables are mapped to their actual phonetic equivalence for pronunciation. Utilizing this data, voice output of unknown words is generated. The actual training of the digital voice library may be conducted manually or by utilizing a neural network. Specification, page 28, lines 13-15.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

1. Whether claims 1-4 are anticipated by Page (U.S. Patent No. 6,175,821).
2. Whether claim 5 is obvious over Page in view of Karaali (U.S. Patent No. 5,668,926).
3. Whether claim 6 is obvious over Page in view of Walker (U.S. Patent No. 6,510,413).
4. Whether claims 7-10 are obvious over Page in view of Lin (U.S. Patent No. 6,076,060).

5. Whether claim 11 is obvious over Page in view Carter (U.S. Patent No. 6,600,814).

VII. ARGUMENT

1. Claims 1-4 (Page)

a. Claim 1

Claim 1 recites a method for converting text to concatenated voice by utilizing a digital voice library and a set of playback rules. The digital voice library includes a plurality of speech items including words and syllables. The digital voice library further includes a corresponding plurality of voice recordings. Each speech item corresponds to at least one available voice recording. The method comprises training the digital voice library to associate each syllable speech item with a literal text syllable of the particular syllable speech item. It is to be appreciated that the digital voice library associates each syllable speech item with a literal text syllable of the particular syllable speech item.

This is exemplified in Figures 6-7. The prior art fails to suggest this specifically recited combination including the association of each syllable speech item with a literal text syllable of the particular syllable speech item.

Page does describe the generation of voice messages. Page fails to describe or suggest the association of each syllable speech item with a literal text syllable. To properly reject claim 1 under principals of inherency, Page must necessarily incorporate each recited claim feature. But Page does not necessarily incorporate each recited claimed feature. Page does describe a text to speech synthesizer for converting text into a series of diphones and concatenating waveforms representing each of the diphones together in order to form a

synthesized speech signal which corresponds to the text of the sentence. Nevertheless, it cannot be inherent in Page that the training occurs as recited by claim 1. After all, Figure 2 shows portions U1, U2, and V that contain multiple words (U1 and U2) or a single word (V). There is no suggestion of the association of each text syllable item with a literal text syllable of the particular speech item as recited by claim 1. In contrast, Page, at least in Figure 2, shows the association of words and groups of words. To this extent, Page teaches away from the innovative training technique recited by claim 1, and instead, utilizes a traditional training technique involving words and groups of words. Thus, Page has shortcomings that are only addressed by the claimed invention.

Thus, Page fails to suggest the recited combination and the concepts of the invention cannot be deemed inherent in Page as Page suggests the use of a traditional training technique as opposed to the approach defined by claim 1. Moreover, Page tends to teach away from the claimed invention, and there is no motivation to modify Page to achieve the claimed invention.

Regarding the final action, although any library training would create some mapping, the claimed association is not suggested by the prior art.

b. Claims 2-4

Claim 2 is believed to be separately patentable from claim 1. Claims 3-4 depend from claim 2. Claim 2 recites receiving a sequence of words including known words that correspond to word speech items in the digital voice library and including unknown words. Each known word is converted into a word speech item in accordance with digital voice library. For each unknown word, the unknown word is parsed to determine a sequence of literal text syllables. The text syllables sequence is converted to a sequence of syllable speech items in accordance with the digital voice library. Claim 2 recites an innovative technique for

handling unknown words in a method for converting text to concatenated voice. The parsing of an unknown word to determine the sequence of literal text syllables, converting the text syllable sequence to a sequence of syllable speech items in accordance with digital voice library, in the recited combination, is not suggested by Page.

The Examiner contends that the invariable and variable portions of the message referred to by Page suggest the subject matter of claim 2. Although Page does mention the formation of synthesized speech that includes variable and invariable portions, there is no suggestion of the parsing of an unknown word to determine a sequence of literal text syllables in combination with the other recited limitations set out in claim 2. After all, Page utilizes a training technique involving words and groups of words, and not literal text syllables.

2. Claim 5 (Page in View of Karalli)

Claim 5 is believed to be patentable due to its dependencies.

3. Claim 6 (Page in View of Walker)

Claim 6 is believed to be patentable due to its dependencies.

4. Claims 7-10 (Page in View of Lin)

Claims 7-10 are believed to be patentable due to their dependencies.

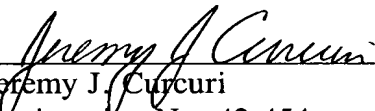
5. Claim 11 (Page in View of Carter)

Claim 11 is believed to be patentable due to its dependency.

The fee of \$330.00 as applicable under the provisions of 37 C.F.R. § 1.17(c) is enclosed. Please charge any additional fee or credit any overpayment in connection with this filing to our Deposit Account No. 02-3978.

Respectfully submitted,

Eliot M. Case

By: 
Jeremy J. Curcuri
Registration No. 42,454
Attorney for Applicant

Date: September 22, 2004

BROOKS KUSHMAN P.C.
1000 Town Center, 22nd Floor
Southfield, MI 48075-1238
Phone: 248-358-4400
Fax: 248-358-3351

Enclosure - Appendix

IX. APPENDIX - CLAIMS ON APPEAL

1. A method for converting text to concatenated voice by utilizing a digital voice library and a set of playback rules, the digital voice library including a plurality of speech items including words and syllables and a corresponding plurality of voice recordings wherein each speech item corresponds to at least one available voice recording, the method comprising:

training the digital voice library to associate each syllable speech item with a literal text syllable of the particular syllable speech item.

2. The method of claim 1 further comprising:

receiving a sequence of words including known words that correspond to word speech items in the digital voice library and including unknown words;

converting each known word into a word speech item in accordance with the digital voice library; and

for each unknown word, parsing the unknown word to determine a sequence of literal text syllables and converting the text syllable sequence to a sequence of syllable speech items in accordance with the digital voice library.

3. The method of claim 2 further comprising:

converting the sequence of word speech items and syllable speech items into a sequence of voice recordings in accordance with the set of playback rules.

4. The method of claim 3 further comprising:

generating voice data based on the sequence of voice recordings by concatenating adjacent recordings in the sequence of voice recordings.

5. The method of claim 4 wherein training the digital voice library further comprises:

utilizing a neural network having an input and an output to train the digital voice library with the neural network receiving the literal text syllable of the particular syllable speech item as input and with the neural network outputting the associated syllable speech item.

6. The method of claim 4 wherein training the digital voice library further comprises:

manually associating each syllable speech item with the literal text syllable of the particular syllable speech item.

7. The method of claim 4 wherein, for each unknown word, parsing and converting further comprises:

parsing the unknown word to determine a sequence of literal text syllables and known words, and converting the sequence to a sequence of syllable speech items and word speech items in accordance with the digital voice library.

8. The method of claim 7 wherein parsing further comprises:
parsing the unknown word in the forward direction to determine any known words;

parsing the unknown word in the reverse direction to determine any known words;

where any known words overlap, selecting the larger word;
parsing the unknown word in the forward direction to determine any literal text syllables; and
parsing the unknown word in the reverse direction to determine any literal text syllables.

9. The method of claim 7 wherein multiple voice recordings that correspond to a single speech item represent various inflections of that single speech item, and wherein converting the sequence of word speech items and syllable speech items further comprises:

determining a desired inflection for each speech item in the sequence of speech items based on the set of playback rules; and

determining a sequence of voice recordings by determining a voice recording for each speech item based on the desired inflection for the particular speech item and based on the available voice recordings that correspond to the particular speech item.

10. The method of claim 7 wherein multiple voice recordings that correspond to a single speech item represent various inflections and ligatures of that single speech item, and wherein converting the sequence of word speech items and syllable speech items further comprises:

determining a desired inflection and desired ligatures for each speech item in the sequence of speech items based on the set of playback rules; and

determining a sequence of voice recordings by determining a voice recording for each speech item based on the desired inflection and desired ligatures for the particular speech item and based on the available voice recordings that correspond to the particular speech item.

11. The method of claim 4 comprising:

for each unknown word, after the unknown word is parsed, storing results of the parsing in the digital voice library so that a next encounter with the same unknown word may be handled more efficiently.